Absenteeism, earnings and biological gender differences

Andrea Ichino
EUROPEAN UNIVERSITY INSTITUTE

Enrico Moretti
UNIVERSITY OF CALIFORNIA, BERKELEY

July 26, 2005

Motivation

Absenteeism is higher among female workers.

This is true even if we consider only illness-related episodes (excluding maternity-related absences).

The cause of this difference is, however, less obvious.

Paringer (1983), Leigh (1983) and Vistnes (1997) consider health as a possible cause, but have no exogenous determinants of illness conditions.

In this paper we show that:

- part of this gender difference in absenteeism may be attributed to a *biological* difference between males and females;
- this biological difference has important earning consequences for females.

Result 1: what is the biological difference?

We use information on the exact date and duration of absence episodes in the personnel dataset of a large Italian bank.

We find that:

- the hazard of an absence due to illness increases for females, relative to males, by 35 percent between the 26th and 30th day after the beginning of a previous episode;

- the spike in the hazard is more evident precisely at 28 days;

- the increase in the hazard is evident only for females younger than 45.

These results suggest that menstrual cycles induce an increase in the hazard of an absence episode.

Result 2: what drives the cyclicality?

The increase in the hazard of an absence due to the 28-days cycle:

- *is higher* among workers in line for merit promotions;

- *does not change* according to average local absenteeism;

These findings suggest that real pain may be an important cause of the cycle.

Result 3: What is the effect on the earnings gap?

28-days cyclical absences explain:

- about 25% of the gender gap in total days lost for absences shorter than 4 days;
- about 79% of the gender gap in the number of absences shorter than 4 days.

In terms of earnings, we find that:

- in our sample, females earn 15% less than males (controlling for age and non-cyclical absenteeism);
- higher absenteeism induced by 28-days cycles explains
  - 3 percentage points or
  - 20 percent

  of this earnings differential.

Result 4: Different "quantities" or different "prices"?

The effect on the earnings gap is mainly a result of:

- different "quantities"
- not of different "prices"

of cyclical absences for males and females.

We discuss possible interpretations of these findings.

Table 1: Selected descriptive statistics

	All		Absent at least once	
	Females	Males	Females	Males
Age	35,7	40.5	35.6	40.3
	(7.8)	(7.8)	(7.9)	(7.8)
Years of schooling	13.3	13.0	13.3	13.0
	(2.7)	(3.3)	(2.7)	(3.3)
Tenure	13.1	16.4	13.0	16.2
	(7.7)	(8.0)	(7.7)	(7.9)
Av. monthly wage (Euros)	1884	2590	1872	2537
	(560)	(1185)	(539)	(1133)
Percent married	63.0	78.6	63.3	78.7
Percent with children	46.5	61.3	47.7	66.9
Percent manager	3.5	19.7	3.3	18.3
Percent high white collar	5.2	11.2	5.1	11.1
Percent low white collar	90.4	64.6	90.7	65.9
Number of observations	3040	13001	2965	11892
Percent of total	19.0	81.0	19.9	80.1

Nature and regulation of absence episodes in the bank

- The analysis is restricted to absence episodes classified as due to illness;
- Workers can have an (almost) unlimited number of illness absences;
- Workers have to send a medical certificate (easy to obtain) if absence longer than 3 days;
- Workers are subject to the possibility of a medical control at home;
- However, this control can occurr only at previously specified times during the day.

Table 2: Distribution of the number of absence episodes by gender

	Females	Males
Mean	11.5	6.6
St Dev.	9.8	7.7
Min	0	0
Max	97	85
Percentile 1	0	0
Percentile 5	1	0
Percentile 10	2	1
Percentile 25	5	2
Median	9	4
Percentile 75	16	8
Percentile 90	24	15
Percentile 95	30	20
Percentile 99	44	38
Number of observations	3040	13001

Table 3: Absenteeism differential by gender controlling for observable characteristics

	i	ii	iii	iv	v	vi
Female - Males	4.645*	4.454*	4.098*	3.519*	4.413*	3.856*
	(0.196)	(0.345)	(0.309)	(0.490)	(0.406)	(0.505)
N. obs:	16041	3672	2017	4283	3258	422
R-sq	0.162	0.138	0.181	0.095	0.179	0.213

Note: The table reports the coefficient (standard errors in parentheses) of the female dummy in regressions in which the dependent variable is the number of absence episodes occurring between January 1, 1993 and February 29, 1996. The sample and the included controls change across columns in the following way:

i. all sample (described in the first two columns of Table 1), controlling for age, tenure, 12 hierarchical level dummies, marital status dummy, southern birth dummy, working in the south dummy, years of schooling and number of children;

ii. not married and without children, controlling for age, tenure, 12 hierarchical level dummies, southern birth dummy, working in the south dummy and years of schooling;

iii. age less than 30, controlling for age, tenure, 12 hierarchical level dummies, marital status dummy, southern birth dummy, working in the south dummy, years of schooling and number of children;

iv. managers, controlling for age, tenure, 4 hierarchical level dummies, marital status dummy, southern birth dummy, working in the south dummy, years of schooling and number of children;

v. college graduates, controlling for age, tenure, 12 hierarchical level dummies, marital status dummy, southern birth dummy, working in the south dummy and number of children;

vi. not married, without children, with age less than 30 and college graduate, controlling for age, tenure, 4 hierarchical level dummies, southern birth dummy, working in the south dummy.

Evidence on Result 1: the biological difference

To identify the existence of a 28-days cycle in absenteeism we follow two alternative approaches:

- Duration analysis:
  - non-parametric hazard estimates;
  - parametric hazard estimates.
- Analysis of the distance between all pairs of (short) absence episodes.

Both approaches generate consistent results.

Duration analysis

We estimate a model for the duration of the spell between the starting dates of two subsequent absence episodes.

The analysis is restricted to the spells starting with the first absence episode occurring between January 1, 1993 and February 29, 1996.

The resulting sample is made of 14857 workers of which 2965 (19.9 percent) are females. (See last two columns of Table 1).

The analysis is further restricted to the first spell following the beginning of the first absence episode: thus we have one spell per worker.

Censoring of spells occurs in 1523 cases.

Non-parametric hazard estimates

Figure 1a - 3b plot, by gender, Kaplan-Meier estimates of the hazard of a second absence episode for durations up to seventy days.

At least four facts are apparent in these figures:

i. As expected, the hazard is almost always higher for females.

ii. For both genders there are spikes at durations equal to 7 or multiples of 7.

iii. At durations equal to 28 and 56 the spike for females is more pronounced than the spike for males.

iv. The spike for females at 28 days is particularly evident only for subjects younger than 45.

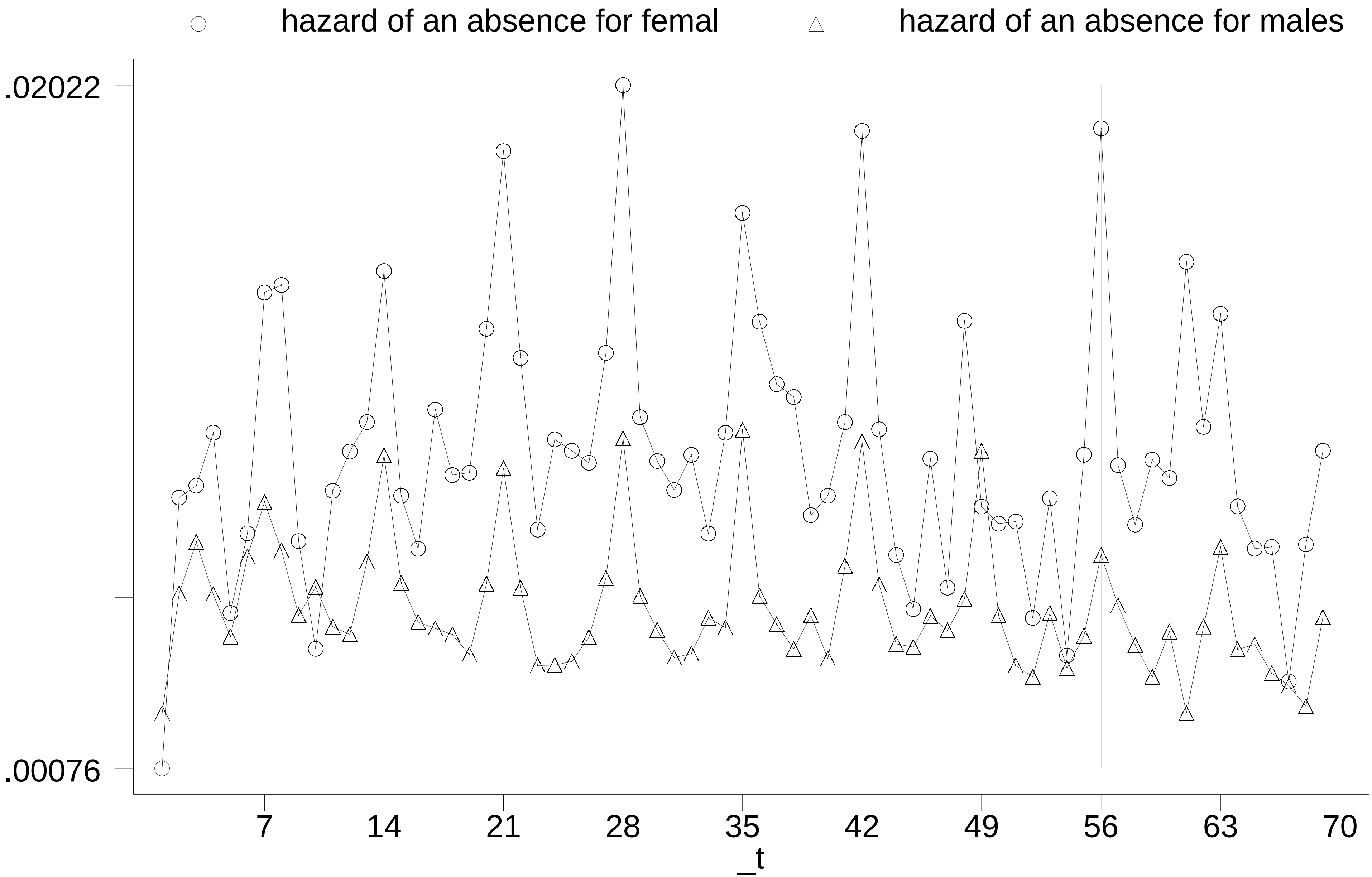

Fig. 1a: female and male hazard rates − all ages

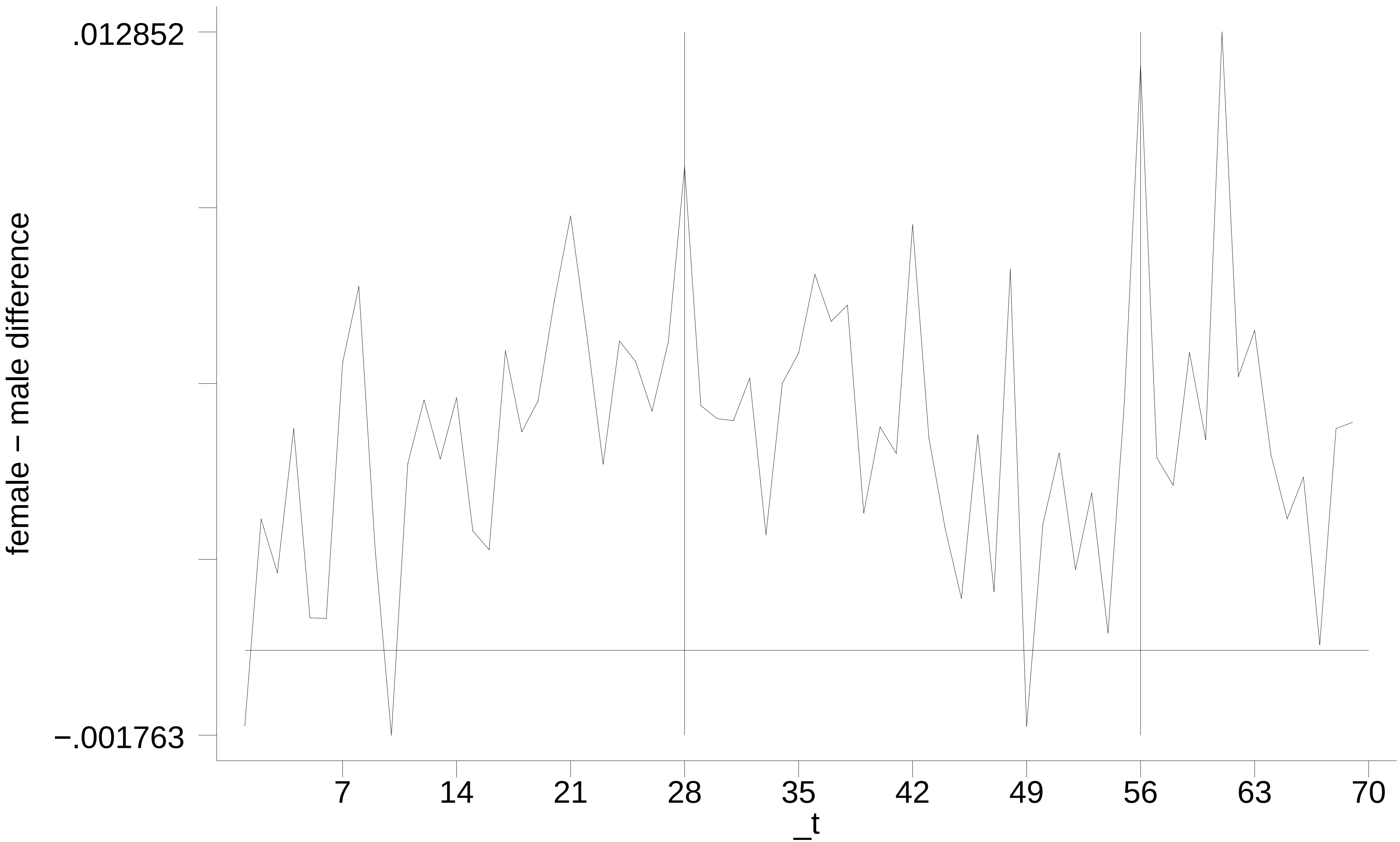

Fig. 1b: difference between female and male hazard rates − all ages

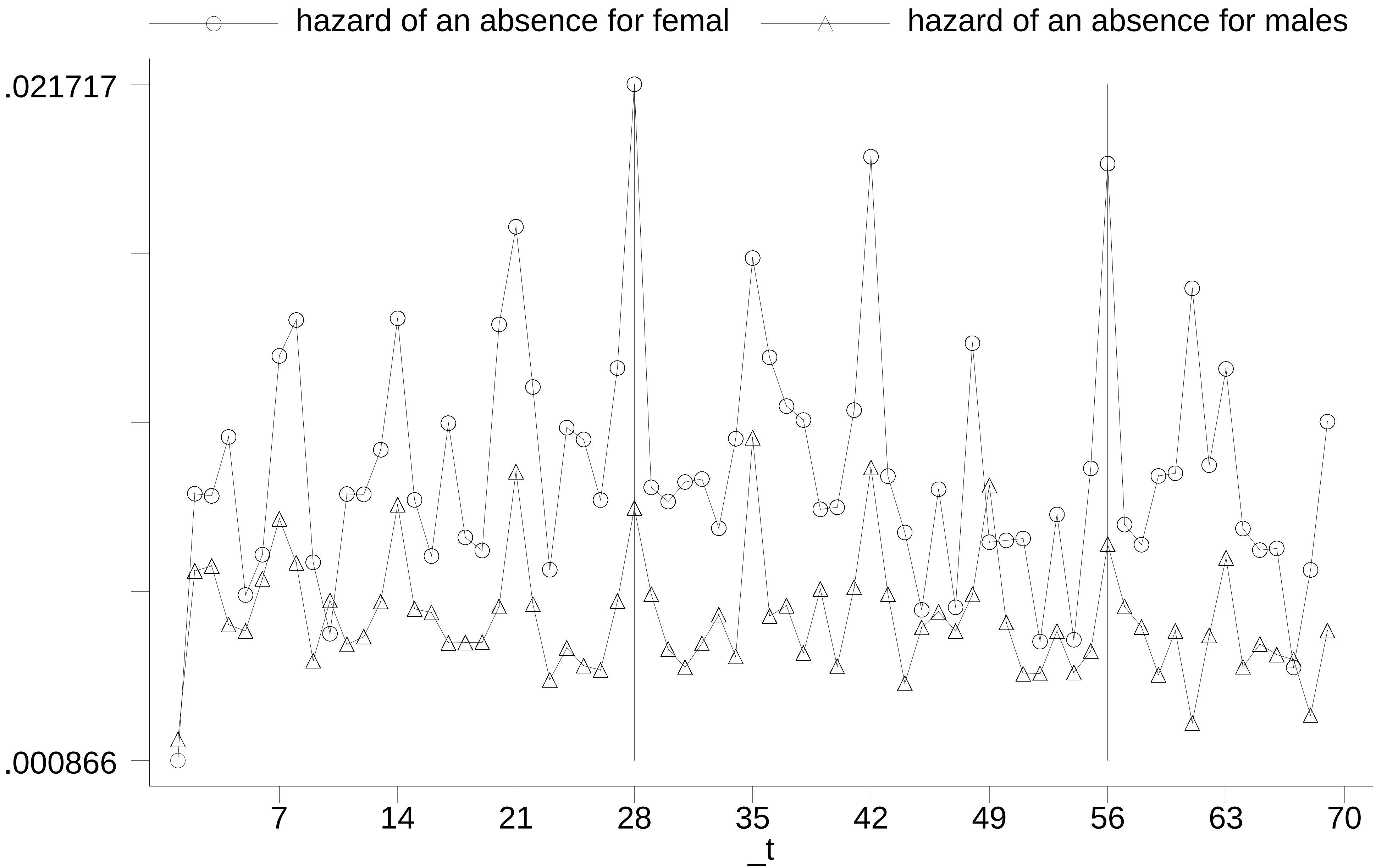

Fig. 2a: female and male hazard rates − under 45

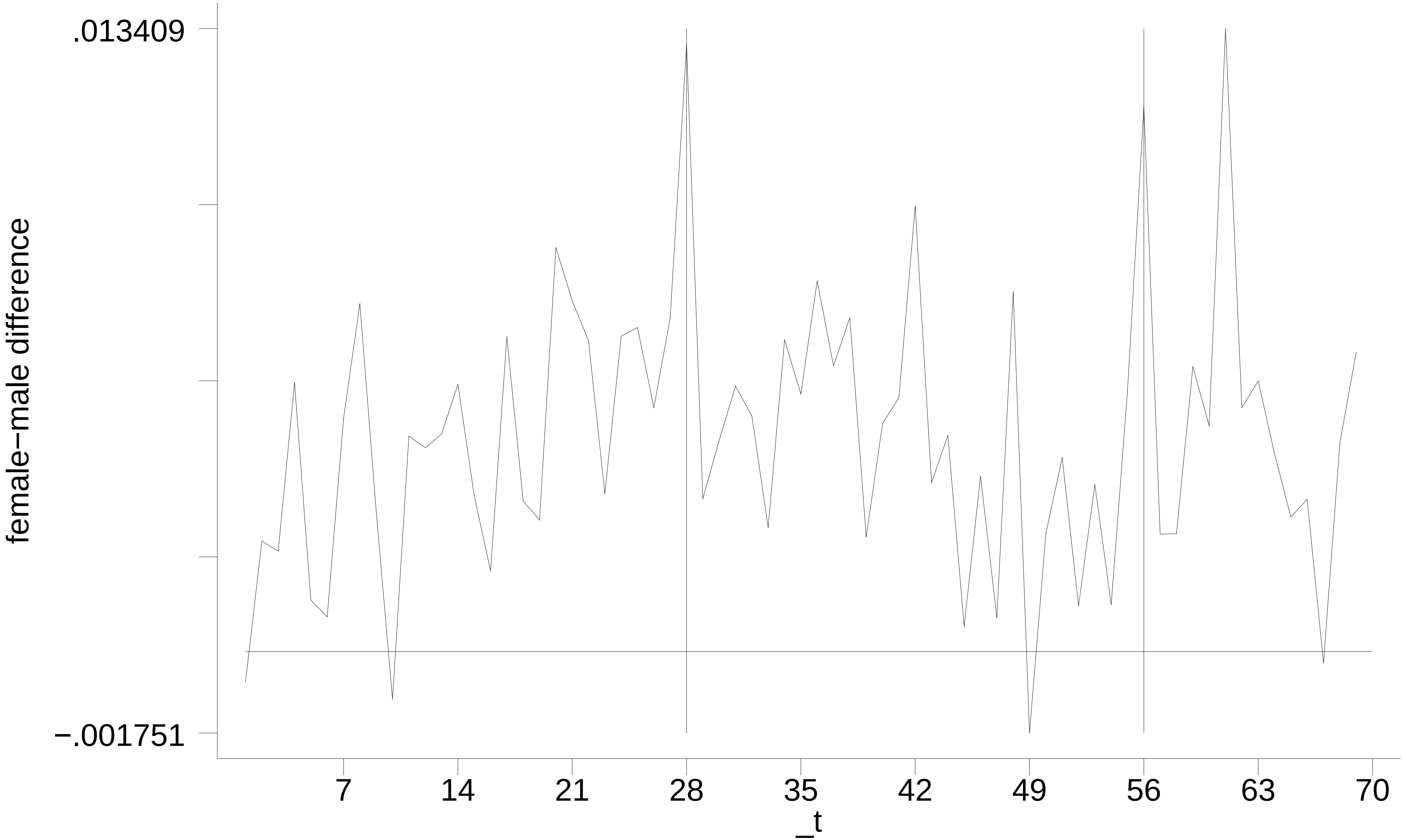

Fig. 2b: difference between female and male hazard rates − under 45

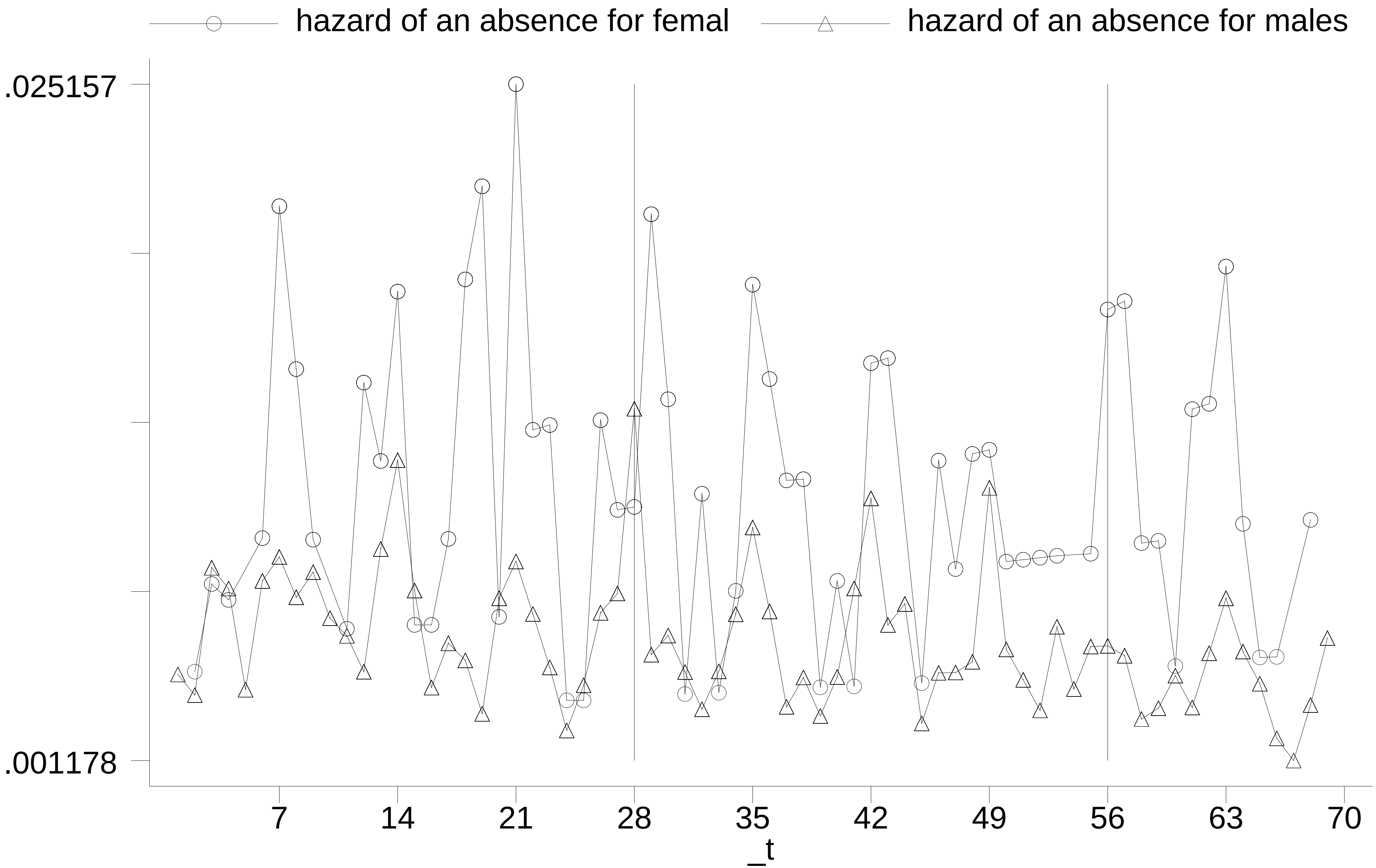

Fig. 3a: female and male hazard rates − above 45

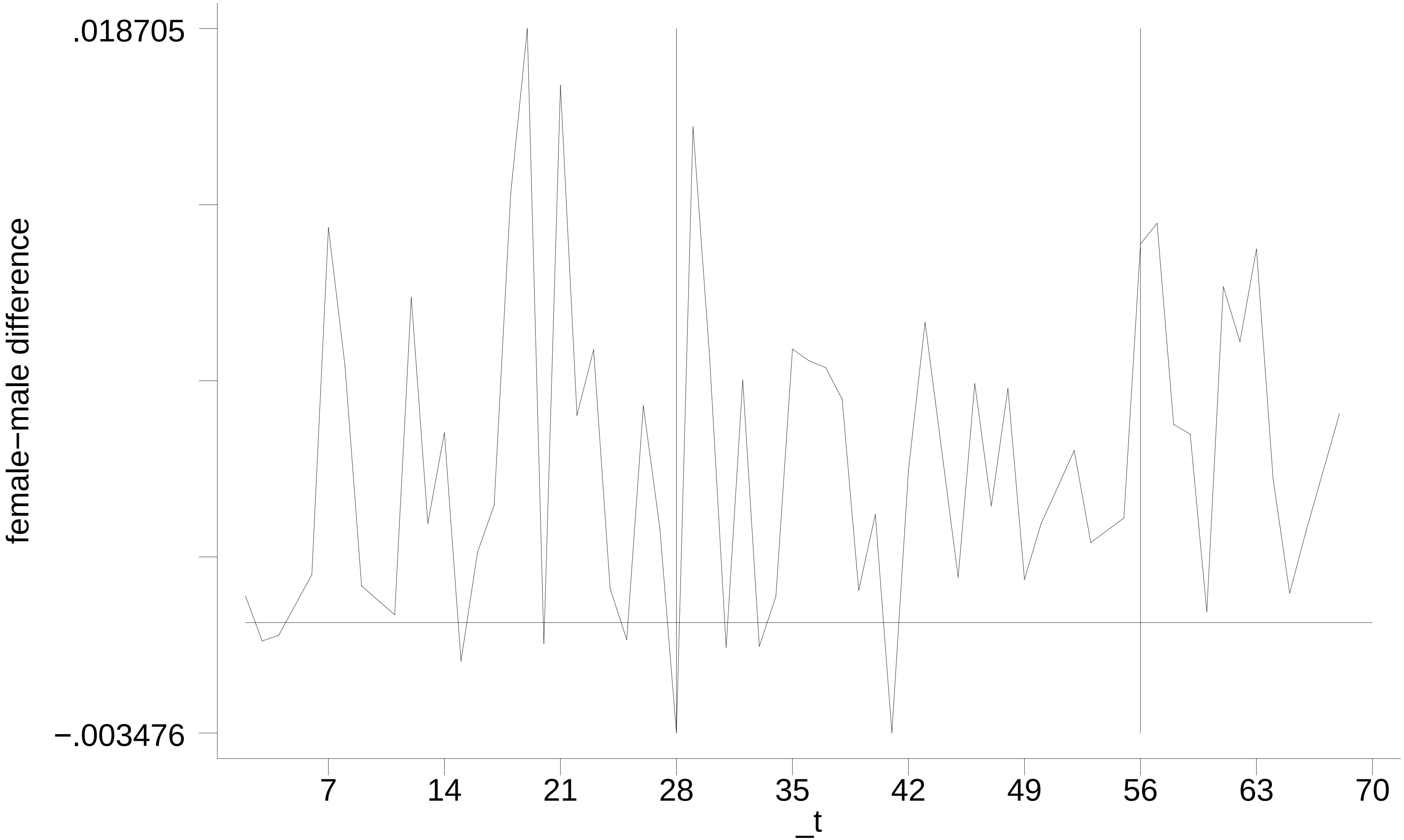

Fig. 3b: difference between female and male hazard rates − above 45

The 7-days periodicity

For both genders,

- the most likely weekday in which an absence begins is Monday;
- the probability that a weekday is the initial date of an absence decreases monotonically between Monday and Friday.

Interestingly, there is a 7-days periodicity independently of the Monday effect, which is probably due to family and other non-work related commitments.

Since 28 is a multiple of 7, the 7-days periodicity creates a confounding pattern with respect to the pattern potentially induced by menstrual cycles.

Parametric estimates allow us to control appropriately for this confounding effect in order to test for the existence of a 28-days periodicity.

Table 4: Distribution by gender of the week-day in which an absence begins.

Week day in which absence begins	Females	Males	Total
Sunday	1	9	10
	(0.03)	(0.08)	(0.07)
Monday	972	4184	5156
	(32.78)	(35.18)	(34.70)
Tuesday	630	2366	2996
	(21.25)	(19.90)	(20.17)
Wednesday	527	1999	2526
	(17.77)	(16.81)	(17.00)
Thursday	482	1872	2354
	(16.26)	(15.74)	(15.84)
Friday	351	1438	1789
	(11.84)	(12.09)	(12.04)
Saturday	2	24	26
	(0.07)	(0.20)	(0.18)
Total	2965	11892	14857
	(100.00)	(100.00)	(100.00)

Parametric hazard estimates

$$h(t, X_{it}, \Psi) = \lambda(t) e^{\alpha + \beta F_i + \gamma M_{it} F_i + \delta S_{it} F_i} \quad (1)$$

where:

- $X_{it} = (F_i, M_{it}, S_{it})$;
- $\Psi = (\alpha, \beta, \gamma, \delta)$;
- $\lambda(t)$ is the baseline hazard, which takes care of the 7-day periodicity;
- $F_i = 1$ indicates that i is a female;
- $M_{it} = 1$ if $t \in [26, 30]$:
  - i.e. whatever the baseline hazard, females have a higher hazard of absence between 26 and 30 days from the previous absence episode;
- $S_{it} = 1$ if $t =$ 7 or multiples of 7:
  - i.e. females are allowed to have a different 7-day periodicity.

Table 5: Hazard of an absence for females relative to males and the risk of a menstrual cycle.

	Days at risk of menstrual cycle: $t \in [26, 30]$		Days at risk of menstrual cycle: $t = 28$		Days at risk of menstrual cycle $t = 28$ or 56	
	e^{β}	e^{γ}	e^{β}	e^{γ}	e^{β}	e^{γ}
All workers	1.58	1.23	1.59	1.15	1.59	1.38
N = 14857	(20.29)	(2.00)	(20.84)	(0.75)	(20.84)	(2.07)
Under 45	1.56	1.35	1.57	1.49	1.57	1.58
N = 10793	(17.89)	(2.57)	(18.47)	(1.89)	(18.47)	(2.65)
Above 45	1.58	1.07	1.58	0.35	1.58	0.66
N = 4064	(7.67)	(0.29)	(7.82)	(-1.67)	(7.82)	(-0.95)
Under 45, controlled	1.56	1.35	1.58	1.49	1.58	1.58
N = 10793	(17.08)	(2.53)	(17.61)	(1.88)	(17.61)	(2.65)
Above 45, controlled	1.42	1.06	1.43	0.35	1.43	0.66
N = 4064	(5.77)	(0.26)	(5.89)	(-1.67)	(5.89)	(-0.95)

“Real pain” or “social norm”?

Two possible reasons for association between menstrual cycles and the hazard of an absence episode:

- menstrual pain may be so strong to induce an unavoidable absence;
- because of a social norm, an absence may be more justifiable for a female if it is associated with menstruations.

We try to discriminate between these hypotheses by:

- restricting the analysis to workers who are more likely to be in line for a promotion based on merit, and therefore less likely to shirk.
- distinguising between environment with different degrees of absenteeism:
  - branches with many females;
  - southern branches.

Findings on “real pain” or “social norm”?

For workers in line for a promotion we show that:

- the hazard of an absence episode for females more than doubles, relative to males, during the days at risk of a menstrual cycle;
- this effect is considerably larger than the one observed for all workers.

In branches with many females or located in the South we show that:

- both these environments are associated with higher absenteeism on average, but
- the hazard of an absence during the days of a potential menstrual cycle is not more evident for females working in these branches.

This evidence supports the hypothesis that the 28-days cycle is mainly caused by the real pain of menstruations.

Table 6: Absenteeism by gender at different ages and hierarchy levels

	Below age 45			
	Bottom levels		Top levels	
	Males	Females	Males	Females
Mean	8.1	12.1	5.1	8.8
Median	6	10	4	7
n. obs.	5602	2205	3213	420

	Above age 45			
	Bottom levels		Top levels	
	Males	Females	Males	Females
Mean	10.4	16.5	5.8	9.2
Median	7	13	4	6
n. obs.	1183	214	1894	126

Table 7: Hazard of an absence for females relative to males and the risk of a menstrual cycle at the top of the firm's hierarchy

Days at risk of menstrual cycle: $t \in [26, 30]$		
	e^{β}	e^{γ}
All workers	1.44	1.35
N = 5653	(7.27)	(1.15)
Under 45	1.42	2.06
N = 3302	(5.93)	(2.46)
Above 45	1.45	0.39
N = 2351	(3.84)	(-1.28)
Under 45, controlled	1.46	2.03
N = 3302	(6.16)	(2.42)
Above 45, controlled	1.34	0.38
N = 2351	(2.90)	(-1.30)

Table 8: The role of different work environments

	Female environment		Southern environment	
	Below 45	Above 45	Below 45	Above 45
Hazard ratio (relative to males) for:				
Females (e^{β})	1.53	1.42	1.57	1.43
	(16.18)	(5.73)	(17.30)	(5.86)
Females in days at risk (e^{γ})	1.56	1.08	1.24	0.91
	(2.07)	(0.15	(1.59)	(-0.30)
Females in days at risk interacted with fraction of females in branch (e^{ϕ})	0.52	0.92		
	(-0.81)	(-0.04)		
Fraction of females in branch (e^{ψ})	1.52	0.99		
	(4.21)	(-0.05)		
Females in days at risk interacted with working in the south (e^{ϕ})			1.29	1.56
			(1.36	(0.32)
Working in the south (e^{ψ})			1.27	1.35
			(10.59)	(7.91)
Controls	yes	yes	yes	yes
Number of workers	10793	4064	10793	4064

What have we learned so far?

- Absences of young females are more likely to have a cycle of 28 days.
- This is not true for females older than 45.
- In the absence of alternative explanations, we conclude that menstruations induce cyclical absences for females under 45.
- The real pain of menstruations (more than a social norm) seems to be the factor that induces cyclical absences.

How much of the gender difference in earnings can be explained by this biological difference?

To answer this question, the first step is to estimate for each worker what fraction of her/his absences is 28-days cyclical.

Methodology to identify cyclical absence episodes

For each worker, we want to estimate what fraction of absences has a cycle of (approximately) 28 days.

We consider the distance between all pairs of short absences (3 days or less).

We call two absences cyclical if their distance is between 26 and 30 days or multiples.

We then normalize this number by the number of short absences so that we obtain an index ranging between 0 and 1.

Note that even for men this indicator may be larger than 0 because of:

- false positive;
- 7-days cycle.

What fraction of absenteeism is due to cyclical episodes?

Table 9: Fraction of absenteeism due to cyclical episodes, by gender - Age under 45

	Males	Females
Ratio of cyclical episodes to short episodes	43.5	60.4

Table 10: Distribution of the fraction of absenteeism due to cyclical episodes, by Gender - Age Under 45

	Males	Females
Share in bottom quantile (low cyclicality)	40.0	18.5
Share in middle quantile	31.1	33.9
Share in top quantile (high cyclicality)	28.8	47.5

Table 11: Number of cyclical episodes in a year, by Gender - Age under 45

Number of cyclical episodes	% Frequency for males	% Frequency for females
0	54	27
1	24	25
2	9	15
3	4	10
4	2	7
5	1	5
6	0	3
7	0	2
8	0	2
9	0	0
10+	0	2

Average differences in days of cyclical absenteeism

The gender difference in the total number of days of absence in a year (conditional on age) is 4.1 (.32).

The gender difference in the number of days lost because of cyclical absence in a year (conditional on age) is 1.0 (.04).

Thus, cyclical episodes are responsible for 25% of the gender gap in absenteeism.

Moreover, the average gender difference masks larger differences in the distributions.

The distribution of cyclical absences for females stochastically dominates the distribution for men.

How much of the gender gap in earnings is explained by cyclical absenteeism?

Suppose that workers can be divided in three groups according to their number of cyclical absence episodes: low, medium high.

Men earnings can be written as

$$Y_m = \pi_{1m} Y_{1m} + \pi_{2m} Y_{2m} + \pi_{3m} Y_{3m} \tag{2}$$

Similarly, women earnings are

$$Y_f = \pi_{1f} Y_{1f} + \pi_{2f} Y_{2f} + \pi_{3f} Y_{3f} \tag{3}$$

The observed and counterfactual earning gaps

The observed earnings gap is therefore

$$Y_m - Y_f = (\pi_{1m}Y_{1m} - \pi_{1f}Y_{1f}) + (\pi_{2m}Y_{2m} - \pi_{2f}Y_{2f}) + (\pi_{3m}Y_{3m} - \pi_{3f}Y_{3f}) \tag{4}$$

We can define a *counterfactual earnings gap*, in the case of no gender differences in cyclical absenteeism, by assigning the distribution of males to females:

$$\tilde{Y}_m - \tilde{Y}_f = \pi_{1m}(Y_{1m} - Y_{1f}) + \pi_{2m}(Y_{2m} - Y_{2f}) + \pi_{3m}(Y_{3m} - Y_{3f}) \tag{5}$$

The difference between the observed and the counterfactual gap can be attributed to menstrual cycles, if these are the only cause of the difference in the π's.

Estimation of earning differences between males and females

In the following model the parameters γ_j estimate the earnings differences $(Y_{jm} - Y_{jf})$ between males and females *within each quantile* j:

$$
\begin{aligned}
logY_i = {} & \beta_1 + \beta_2 C_{2i} + \beta_3 C_{3i} + \\
& \gamma_1 C_{1i} F_i + \gamma_2 C_{2i} F_i + \gamma_3 C_{3i} F_i + \\
& \delta_2 N_{2i} + \delta_3 N_{3i} + \mu X_i + e_i
\end{aligned} \tag{6}
$$

where:

- C_{Ji} is an indicator for the j quantile of the cyclical absences distribution;
- N_{Ji} is an indicator for the j quantile of the non-cyclical absences distribution;
- F_i is an indicator for females;
- X_i is a quadratic polynomial in age.

This is essentially a "diff. in diff." estimator in which males are used as the counterfactual for females in the absence of menstrual cycles.

Estimates

Table 12: Earnings Equation

	(1)	(2)	(3)
Female	-.209	-.151	-
	(.006)	(.006)	
Medium number of cyclical absences			-.052
			(.006)
High number of cyclical absences			-.132
			(.007)
Small number of cyclical absences * female			-.141
			(.012)
Medium number of cyclical absences * female			-.140
			(.010)
High number of cyclical absences * female			-111
			(.009)
Controls for non-cyclical absences	N	Y	Y
Controls for age	N	Y	Y

Main finding

The bottom line is:

- a cyclical absence "costs" approximately the same for males and females;
- but females have more cyclical absences than males (because of menstrual cycles).

The total gender gap in earnings between males and females is 15%

The gender gap if males and females had the male distribution of cyclicality would be 12%

Under our assumptions we conclude that if females did not have menstruations, the gender gap in earnings would be 20% lower.